



Stochastic Variation: A Comment on Lewis-Beck and Skalaban's "The R-Square"

Citation

King, Gary. 1991. Stochastic variation: A comment on Lewis-Beck and Skalaban's "The R-Square". *Political Analysis* 2(1): 185-200.

Published Version

<http://pan.oxfordjournals.org/>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4313306>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Stochastic Variation: A Comment on Lewis-Beck and Skalaban's "The R -Squared"

Gary King

1. Introduction

In an interesting and provocative article, Michael Lewis-Beck and Andrew Skalaban make an important contribution by emphasizing several philosophical issues in political methodology that have received too little attention from methodologists and quantitative researchers. These issues involve the role of systematic, and especially stochastic, variation in statistical models. After briefly discussing a few points of disagreement, hoping to reduce them to points of clarification, I turn to the philosophical issues. Examples with real data follow.

2. Does R^2 Contain New Information?

Suppose we run a linear regression and obtain estimates of the effect parameter vector b , the variance matrix $\hat{V}(b)$, and the variance of the disturbances $\hat{\sigma}^2$. If we discard the original data, would we learn something new about a substantive research question by also knowing R^2 ? I argued in my 1986 article that the answer to this question is "no."

To make this point somewhat more vivid, consider the following simple proof. First, let $S(z)$ be the sample variance of some variable z :

$$S(z) = \frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n}, \quad (1)$$

where n is the number of observations and \bar{z} is the sample mean of z .¹ The R^2 -

My thanks goes to Neal Beck and Andrew Gelman for many helpful comments and to the National Science Foundation for grant SES-89-09201.

1. One frequently divides by $n - 1$ instead of n to produce an unbiased estimate, but I use n for simplicity. $S(z)$ also happens to be the maximum likelihood estimate.

statistic is then simply expressed as the ratio of the variance in the fitted values to the variance in the observed values,

$$R^2 = \frac{S(\hat{y})}{S(y)}. \quad (2)$$

Now let us compare the information in this statistic to that in the estimator for the variance of the disturbances,

$$\hat{\sigma}^2 = S(y) - S(\hat{y}), \quad (3)$$

and the variance matrix of the coefficients,

$$\hat{V}(b) = [S(y) - S(\hat{y})](X'X)^{-1}. \quad (4)$$

In comparing equation 2 with equations 3 and 4, it should be obvious that R^2 , $\hat{\sigma}^2$, and $\hat{V}(b)$ contain precisely the same two key pieces of statistical information— $S(y)$ and $S(\hat{y})$. The distinction is only that R^2 expresses them as a ratio, $\hat{\sigma}^2$ expresses them as a difference, and $\hat{V}(b)$ expresses them as a difference, weighted by $(X'X)^{-1}$. Thus, no new information (and certainly no new substantive information) is added by knowing the value of the R^2 -statistic beyond $\hat{\sigma}^2$.²

3. How Should One Express the Information in R^2 ?

We take as given, then, that R^2 and $\hat{\sigma}^2$ (or $\hat{V}(b)$) contain exactly the same variance information. If we know one, we have little reason to determine the other. Nevertheless, we should still consider how best to express this information. While largely a matter of taste, the choice can be important in leading or misleading researchers conducting routine empirical analyses. In this section, I provide three important reasons to prefer the information expressed as the square root of the variance of the disturbances—the standard error of the regression, $\hat{\sigma}$.

First, Lewis-Beck and Skalaban prefer R^2 because it is *standardized* to be

2. Is R^2 an estimate of a population parameter that is a part of the underlying statistical model, as suggested by Lewis-Beck and Skalaban? To answer this question, see equation 1 in their article, a perfectly good statement of a regression model. Note that they have no population parameter on that page (or any other) that corresponds to the sample R^2 . In addition, they make no distributional assumptions when stating this model. Univariate Normality is helpful, if appropriate, but unnecessary nevertheless; multivariate normality, which they mention later on, is not appropriate and would not be helpful even in the extremely unlikely case in which it were satisfied.

bounded between zero and one, whereas $\hat{\sigma}$ is the same information in *unstandardized* form (only bounded below by zero). It is true that R^2 always falls in the same range, but this range has different meanings for different research problems. Furthermore, the unstandardized form is in precisely the units of the original dependent variable. One need not know any statistics to understand that Lewis-Beck and Skalaban's predictions of New Jersey mayoral elections are accurate to within roughly plus or minus 7.7 percentage points (see Lewis-Beck and Skalaban, equation 6). When reporting public opinion poll results, even Dan Rather now routinely gives the 95 percent confidence interval (popularly referred to as the "margin of error"). In contrast, the value of R^2 , 0.52, cannot be interpreted in the same language as one would normally speak about New Jersey elections.

For example, imagine advising the mayor of Newark during his reelection campaign. If your model predicts that he will receive 53 percent of the vote, plus or minus a margin of error of 1 percent, you should have little trouble conveying this information: You are convinced that he will win the election. However, if you were to give him your 53 percent vote prediction with an R^2 value of 0.85 his first question would likely be: "What does this have to do with the outcome of the election?" or "How sure are you of these results?" Indicating to him that you have explained 85 percent of the variance is unlikely to clarify these matters.

Second, we seem to learn again and again in many areas of statistics that standardization can cause trouble. For example, political methodologists now generally recognize that unstandardized regression coefficients should be used instead of standardized coefficients. (Indeed, this example is more than an approximate analogy because a squared standardized regression coefficient is equal to R^2 in the case of a single explanatory variable. Thus, any argument that applies to standardized and unstandardized coefficients should apply with equal weight to standardized and unstandardized variability.) Standardized coefficients are not only farther from the substance of the research problem, but they cause serious problems when comparing results across different data sets. Indeed, this point was made persuasively over a decade ago in a widely read Sage green monograph written by Michael Lewis-Beck (1980, 66): "In general, when the variance in X diverges from one sample to the next, it is preferable to base any cross-sample comparisons of effect on the unstandardized partial slopes." This should provide sufficient reason to prefer the unstandardized $\hat{\sigma}$ to the standardized R^2 -statistic.³

3. Obviously, if you use standardized coefficients or standardized variability and also report the sample variances of your variables, you *could* reach appropriate inferences. However, few methodologists worry about the reader or analyst who would take the time to make such a calculation. What we seek are ways of reporting statistical results that minimize the possibilities for misunderstanding.

Finally, I turn to a more compelling philosophical reason to prefer $\hat{\sigma}$. We all agree that R^2 is a fine measure when one is comparing different specifications with the same dependent variable. Quite commonly, analysts gradually add explanatory variables to their equation and use increases in R^2 as one piece of information in deciding whether the variable should remain in the equation; none of us have any objection to this procedure, even when the comparison starts out with just the constant term in the regression and an R^2 of zero. Measures of variability can also help to answer some important substantive questions, a topic to which I turn in section 5. Our disagreement is over whether R^2 or any measure can provide a way to judge the extent of the discrepancies between the current model and the true model.

Let us imagine the best of all worlds. Suppose one had a way to calculate the probability that a chosen model is the true one, given the data: $\Pr(\text{model}=\text{truth}|\text{data})$. If this quantity, usually called an "inverse probability," were possible to calculate, we could easily use it to choose a model. We would merely calculate the inverse probability for many models and choose the single model with the highest probability of being true. We would then have (1) the single model with the highest probability of being true, given the data, and (2) a measure of how good the chosen model is. When comparing R^2 across equations, or evaluating a single R^2 value and trying to determine how good the model is, people frequently treat this statistic as if it were an inverse probability. They sometimes try to increase R^2 , thinking that a higher R^2 means that you have a better model. Others use language and logic that seems to approximate inverse probability, such as when Lewis-Beck and Skalaban ask "How good is [the model], if it is not perfect?" Similarly, some researchers run a single regression and comment on how impressed they are that R^2 takes on a given value.

The goal of calculating an inverse probability is a worthy one. Indeed, numerous statisticians have made it their life's work during the last three centuries (see Stigler 1986). Unfortunately, scholars eventually realized that this goal was unreachable, even in theory. In the 1920s, R. A. Fisher first saw the impossibility of inverse probability and proposed a practical solution. He separated the two goals of inverse probability and concluded that it was the second that was especially problematic: one could never provide a measure of how close a model approximates the true one. He then introduced the concept of "likelihood" as a way to choose among different models for a given data set. The idea was to choose the model that maximized the likelihood of having generated the data we observe.

Thus, likelihood differs from inverse probability in two important ways. First, nowhere in the definition of likelihood is "truth" mentioned, and there is no way to summarize how close one's model is to reality. Second, the concept of likelihood is entirely relative to one's data. Fisher provided a way of calculating the likelihood of various models having generated one's data and

of choosing the single model that was most likely to have generated it. But all comparisons among the likelihoods of different models must be for a single, given data set. No absolute assessments of the model should or even could be made. As such, it is well known that one should never compare likelihood values across data sets (see Edwards 1972; King 1989).

When using the methods of likelihood, this all becomes obvious since the likelihood function is an unknown function of the data. This makes standardization for the purpose of comparing across data sets impossible: One can never turn a likelihood into an inverse probability.

Indeed, my comparison here between R^2 , likelihood, and inverse probability is more than an analogy. The likelihood in linear-Normal regression models is a mathematical function of R^2 , or $\hat{\sigma}$ (King 1989, 85, n. 21). Because likelihoods provide no way to make an absolute assessment of how close a model is to reality, R^2 cannot provide a way either. Perhaps because the scale of R^2 is fixed from zero to one, it *appears* comparable across data sets, but this is simply not true.

4. Over- and Underfitting

The estimate $\hat{\sigma}$ dominates R^2 in virtually every way, but if one is careful to report sample variances, and to interpret results properly, one could use R^2 without being misled. As noted above, it is a statistic based on precisely the same information as in $\hat{\sigma}$.

Suppose now that Lewis-Beck and Skalaban were to agree to dismiss the use of R^2 in cases where one seeks to measure the discrepancy between a chosen model and the true one, and when comparing R^2 across regressions based on different dependent variables. Nevertheless, R^2 (or $\hat{\sigma}$) still can be useful in comparing specifications for a single data set, a topic to which I now turn.

To understand these issues, I distinguish two fundamental views of random variation.⁴ These two perspectives are extremes on a continuum. Although significant numbers of scholars can be found who are comfortable at the extremes, most political scientists fall somewhere between the two. (I find that economists tend to be closer to perspective 1, whereas statisticians are closer to perspective 2.⁵)

Perspective 1: Random variation exists in nature. Even if we measured all variables without error, collected a census of data, and included every conceivable explanatory variable, our regressions would never predict

4. I have never seen these views clearly delineated, but the difference among scholarly perspectives is quite pronounced.

5. Perspective 1 is also very common in the field of engineering called "quality control."

exactly. A researcher can divide the world into apparently systematic and apparently stochastic components, but nothing a researcher does to analyze data can have any effect on reducing the fundamental amount of stochastic variation existing in various parts of the empirical world.

Perspective 2: Random variation is only that portion of the world for which we have no explanation. The division between systematic and stochastic variation is *imposed* by the analyst and depends on what explanatory variables are available and included in a regression analysis. Given the right explanatory variables, the world is entirely predictable.

Consider now what these perspectives on stochastic variation have to say about choosing regression specifications. Under perspective 1, R^2 has an *unknown* optimal value. If R^2 is below this optimum, the specification is defective because relevant explanatory variables have been omitted: some systematic variation has been incorrectly included in the stochastic component of the statistical model. If R^2 is larger than the unknown optimum value, the specification is defective because of overfitting: irrelevant explanatory variables have been included and the regression merely maps the random error, making the model useless especially for out-of-sample predictions. The problem is that $R^2 = .1$ could be too high, and $R^2 = .9$ could be too low; the researcher can never know. From this perspective, R^2 is very important, but we have no way to use it to assess the model.

Under perspective 2, R^2 is a description of the data. It says nothing of interest either about nature or the parameters we are estimating. $1 - R^2$ represents the variation to be accounted for by the omitted variables, which we might now have the impetus to collect. The concept of overfitting is more difficult to deal with under perspective 2, and, indeed, some adherents take the extreme position that overfitting either is a meaningless concept or has no pernicious effect. This version of perspective 2 is problematic in that the very concept of inference can be overshadowed by mere description of the data.

How then do we decide on whether we are under- or overfitting or if we have a good specification, given a particular dependent variable? Much research has addressed this question, but I limit my discussion to four points that are particularly relevant to R^2 and $\hat{\sigma}$.

First, one purpose of science is to explain many things with a few things. Given a dependent variable and an interesting set of hypotheses, we should prefer to reduce stochastic variation in a model by using as few explanatory variables as possible. Even if one can justify including 25 explanatory variables on theoretical and empirical grounds in a regression, it will often make little sense to include them. The purpose is never to include all explanatory variables that seem important or even have significant t -statistics.

Second, even if one had good reasons for including a large number of

explanatory variables, there is usually an efficiency loss when we estimate additional parameters. Adding variables has a cost in terms of the precision of estimation of the other parameters in the equation, and this cost is not accounted for in R^2 (or even the adjusted R^2 statistics).

Third, we can sometimes learn what our stochastic variation should look like even before beginning data analyses. For example, the stochastic variation produced by random sampling is easily calculable. In addition, we can often anticipate fairly accurately how much measurement error our dependent variable contains, either from intercoder reliability or simply impressions obtained when doing coding. An intelligent analysis of these issues can make statistics like R^2 or $\hat{\sigma}$ much more useful.

For example, the variation across surveys in the marginal percent answering "yes" to a survey question is the total of (1) measurement error, (2) sampling error, and (3) true opinion change. This means that the total of measurement and sampling error *must* be less than or equal to the observed change. If we then did a regression analysis, and $\hat{\sigma}$ was smaller than our estimate of sampling plus measurement error, we know we are overfitting. If this comparison indicates that we might be underfitting, this has very interesting substantive implications. One could learn about either the nature of stochastic variability in this research problem (perspective 1) or the variation to be explained by omitted variables (perspective 2).

One final, but very important, point about choosing a specification has arisen in the econometric forecasting literature. Perhaps the central lesson of this literature is that out-of-sample forecasts are much worse with overfitting than with underfitting. Models with fewer parameters, and thus *lower* R^2 -values, usually forecast *more* accurately. This provides another powerful reason to consider a more intelligent approach to understanding stochastic variation.

5. Examples

I suspect that one of the lasting contributions of Lewis-Beck and Skalaban's article will be their implicit attempt to expand the focus of empirical analyses: point estimates are important but by no means everything we wish to know about the world. In the end, we seek to estimate entire probability distributions, not merely means. The mean of a probability distribution is a point we can (and should) estimate, but it is not a lot more interesting than all the other points comprising the distribution.

As an example, I took data from the 338 districts that held contested elections for the U.S. House of Representatives in both 1984 and 1986.⁶

6. These data are fully analyzed for various purposes in Gelman and King 1990 and King and Gelman 1991.

TABLE 1. Regression Models, 1982-84

Explanatory Variables	Model 1		Model 2		Model 3	
	<i>b</i>	SE	<i>b</i>	SE	<i>b</i>	SE
Constant	-.013	.016	.150	.024	.156	.022
V_1	.957	.028	.636	.045	.618	.041
<i>P</i>			.064	.007	-.041	.015
<i>I</i>					.116	.015
$\hat{\sigma}$.083		.075		.069	
R^2	.782		.822		.849	

Note: $N = 338$

Model 1 in table 1 presents the regression of the Democratic proportion of the two-party vote in 1984 (V_2) on the proportion in 1982 (V_1). The regression line is plotted in figure 1. A large proportion of the districts fall within plus or minus $\hat{\sigma} = 0.08$ of the regression line, an obviously interesting number easily explainable to a nonstatistician (0.08 is eight percent of the vote for congress). The R^2 is .782, but this number is not of much use by itself.

Figure 1 shows that the stochastic variability exists mainly in two clumps, one for Democratic and one for Republican winners in 1982. (In a sense, this is just a two-dimensional "Mayhew Histogram," where few dots appear near 0.5 on both axes due to the small number of marginal districts.) This pattern in the residuals is, of course, an omitted variable, which I call *P* and code as -1 for Republican winners in 1982 and 1 for Democratic winners in 1982. Including this variable fits two regression lines with the same slope. The results for this specification can be found in model 2 in table 1; it is portrayed in figure 2 (with a vertical line added for clarity). Note that the variability around each of these lines is slightly smaller ($\hat{\sigma}$ has declined from 0.083 to 0.075 and, equivalently, R^2 has increased from .782 to .822) and more homoskedastic.

The original purpose in analyzing these data was to study incumbency advantage, so I now include a variable for incumbency status (coded -1 for Republican incumbents, 0 for open seats, and 1 for Democratic incumbents), and include it in this regression equation (see model 3, table 1). The standard error of the regression, $\hat{\sigma}$, has again declined, and R^2 has increased, slightly. These results are portrayed in figure 3. Note that dots represent congressional districts with an incumbent, and circles stand for open seats, in 1982. The dashed lines fit Democratic and Republican open seat races, and the solid lines fit the Democratic and Republican incumbent races. The incumbency advantage is the coefficient on *I* in table 1 and is equal to the vertical difference between the dashed and solid lines in figure 3, a substantial 0.116 fraction of the vote.

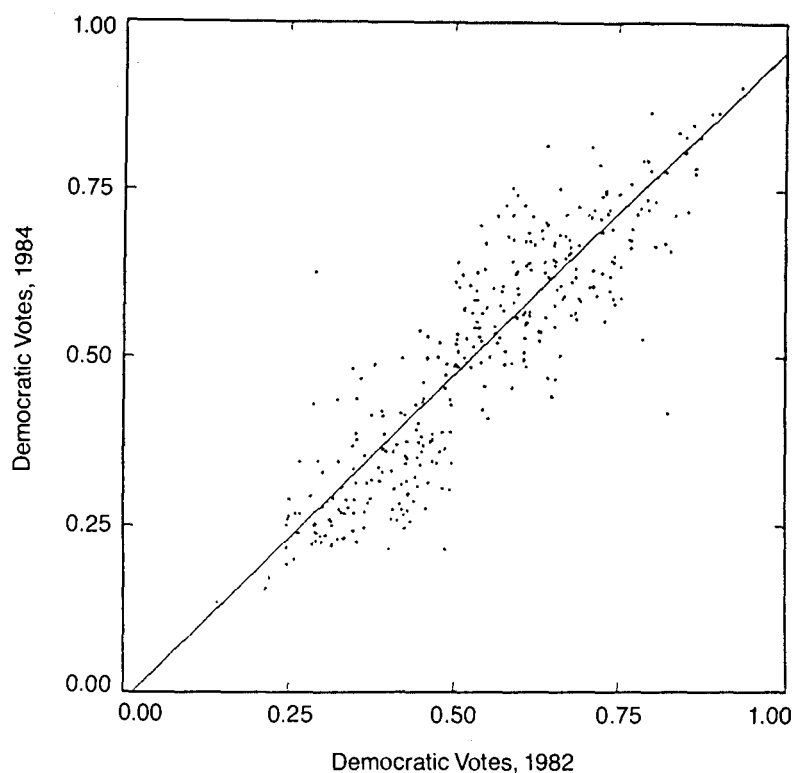


Fig. 1. Electoral swing in contested U.S. House districts

In Gelman and King (1990), we settle on model 3 as the best specification. We also believe that it provides the best estimate of the incumbency advantage, which was the goal of our paper. However, note that all four lines are constrained to have the same slope. A natural question to ask is whether they should be allowed to vary. We will undoubtedly reduce the stochastic variation (increase the R^2) by doing so. At first glance, this seems like a good idea. Model 4 in table 2 includes the interactions necessary to let all the slopes vary. R^2 does increase, and $\hat{\sigma}$ declines, slightly. In addition, an F -test of whether the four extra coefficients are jointly different from zero is significant at the .001 level. Furthermore, figure 4, which plots the results of this regression, indicates that model 4 seems to do quite a good job of reducing the apparent stochastic variability. It also admits a reasonable substantive interpretation: the less competitive a race is in 1982, the larger is the incumbency advantage (the vertical distance between the dashed and solid lines).

But should we include these variables? To test to see whether we have

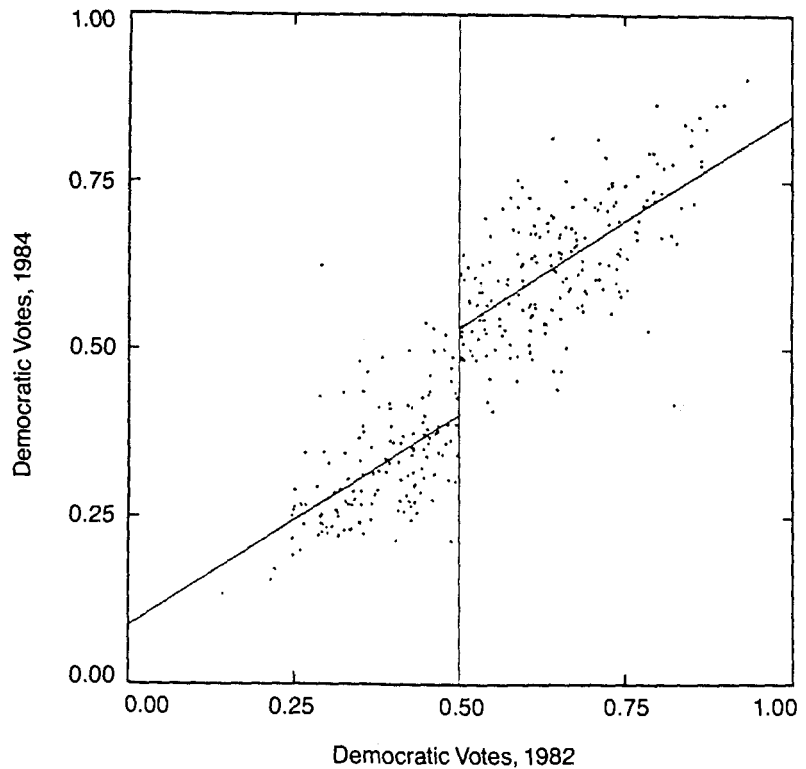


Fig. 2. Electoral swing in contested U.S. House districts with Republican and Democratic regression lines

found a systematic effect, or whether we are just mapping random variability with explanatory variables, it pays to look at another random draw of the same election data. Of course, it is not possible to run all the congressional elections held in 1982 and 1984 over again, but it would be possible to conduct the same analysis for another pair of election years. The only disadvantage of this procedure is the possibility that a fundamental change occurred in the underlying electoral system. I chose the next pair of elections, 1986 and 1988, because of the small probability of such a change.

Figure 5 shows the results of model 4 fit to these new data. Note how the lines, especially the dashed lines, change dramatically from figure 4 to figure 5. This seems to imply that model 4 overfit both set of data and the coefficients on the explanatory variables do not seem to represent real systematic effects. To verify this hypothesis, figure 6 uses the new data to fit model 3. Note how the lines in figures 3 and 6 are almost identical. Even though the

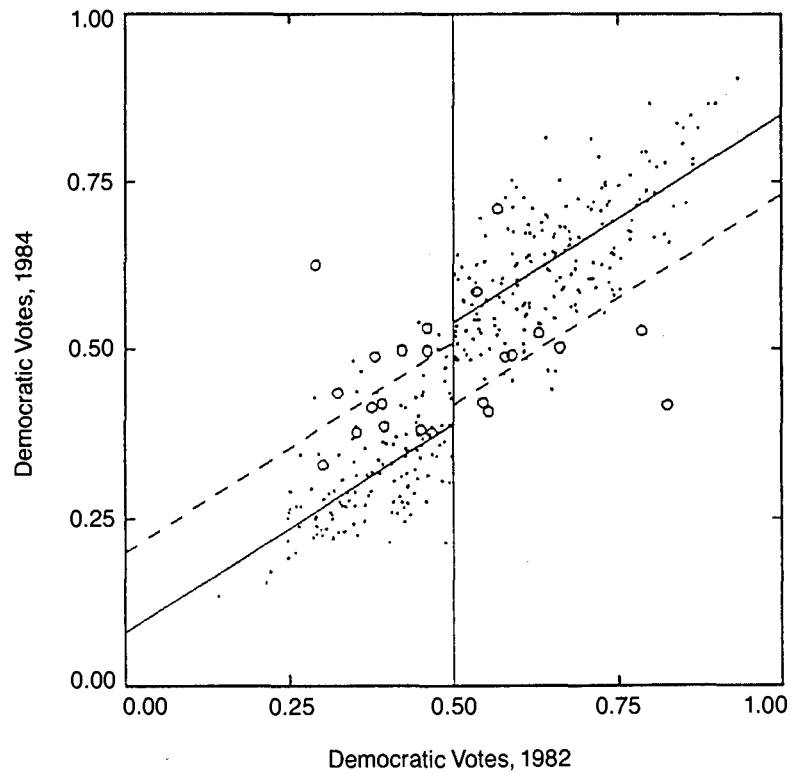


Fig. 3. Electoral swing in contested U.S. House districts. Dots are incumbents and circles are open seats, 1984.

congressional district votes vary around these lines in slightly different ways in the two pairs of election years, model 3 is clearly picking up the systematic component of these congressional elections data. Indeed, although I do not have the room to show it, the same sorts of patterns occur when applying these models to many other election years.

Finally, it is worth studying the several types of variation in these analyses. With all the election years from 1940 to the present, the estimate of incumbency advantage by applying model 3 to each pair of years went from a modest .02 to a very substantial .12. In general, we can always break down this total (or observed) variability by an equation like the following:

$$(\text{Total Standard Deviation})^2 = (\text{Estimation Standard Deviation})^2 + (\text{True Standard Deviation})^2$$

TABLE 2. Standard Regression Model, 1982-84

Explanatory Variables	Model 4	
	<i>b</i>	SE
Constant	.549	.094
<i>V</i> ₁	-.127	.194
<i>P</i>	.080	.094
<i>I</i>	-.018	.097
<i>IV</i> ₁	.079	.199
<i>IP</i>	-.413	.097
<i>PV</i> ₁	-.066	.194
<i>PIV</i> ₁	.780	.199
$\hat{\sigma}$.068	
<i>R</i> ²	.857	

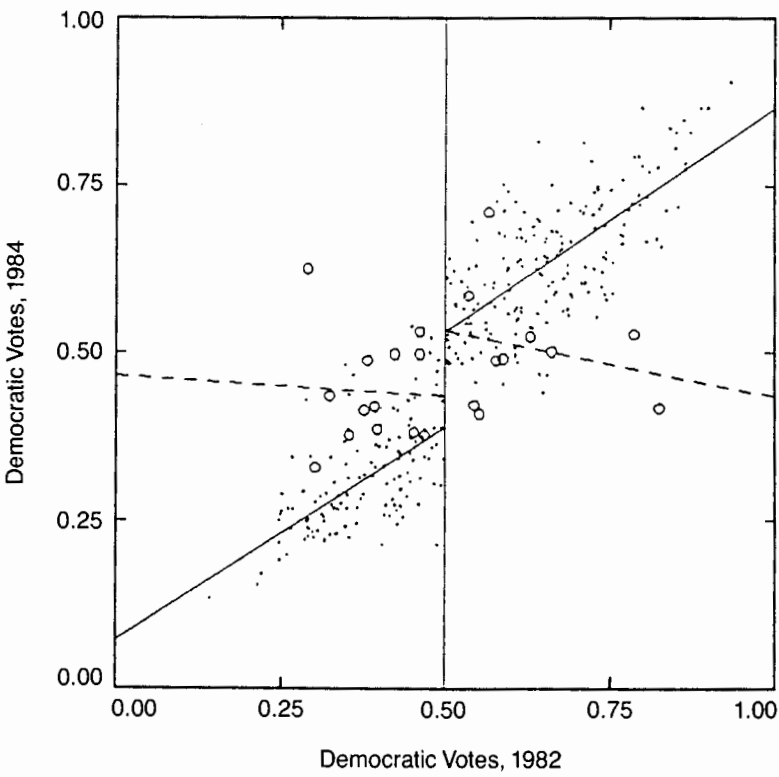


Fig. 4. Electoral swing in contested U.S. House districts. Dots are incumbents and circles are open seats, 1984. Separate slopes and intercepts.

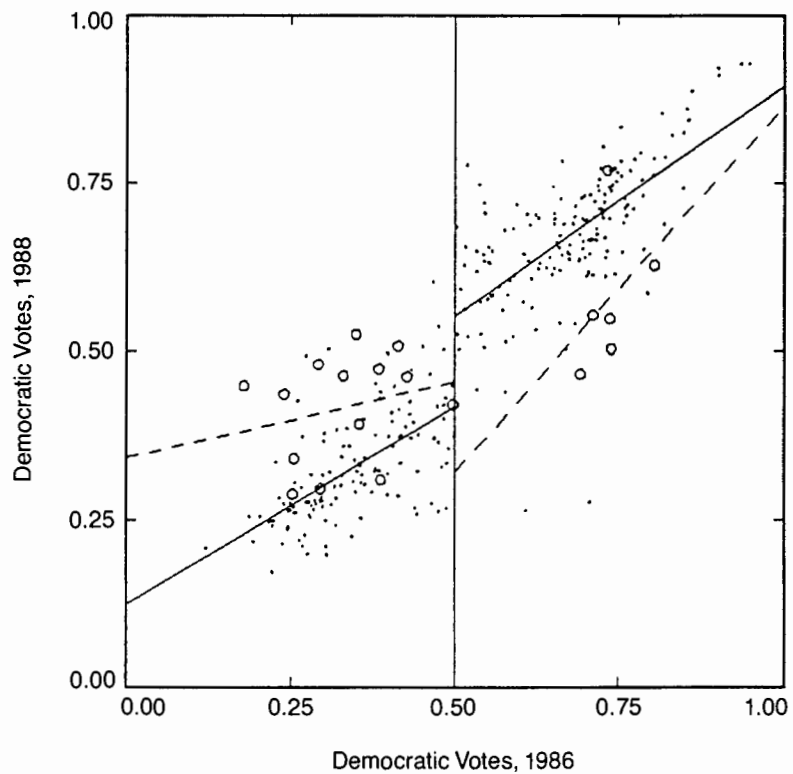


Fig. 5. Electoral swing in contested U.S. House districts. Dots are incumbents and circles are open seats, 1988. Separate slopes and intercepts.

We usually cannot estimate the true standard deviation directly, but it can be inferred by estimating the other two quantities.⁷ For example, the *total standard deviation* in these estimates over the whole period is .043. In addition, the *estimation standard deviation* in the incumbency advantage coefficient (as measured by the average standard error on the incumbency variable across applications of model 3 to successive pairs of election years) is .012.⁸ This estimation variability means that we know the incumbency advantage fairly precisely, especially given its level in recent years (about .12). But this also

7. The square of a standard deviation is a variance, required to make these quantities add. Nevertheless, the standard deviation (the square root of the variance) is more interesting because it is in the units of the dependent variable—in this case, proportions.

8. To fix these ideas, recall that the incumbency advantage estimate in 1982–84 was .116 with a standard error of .015; see the last column of table 1.

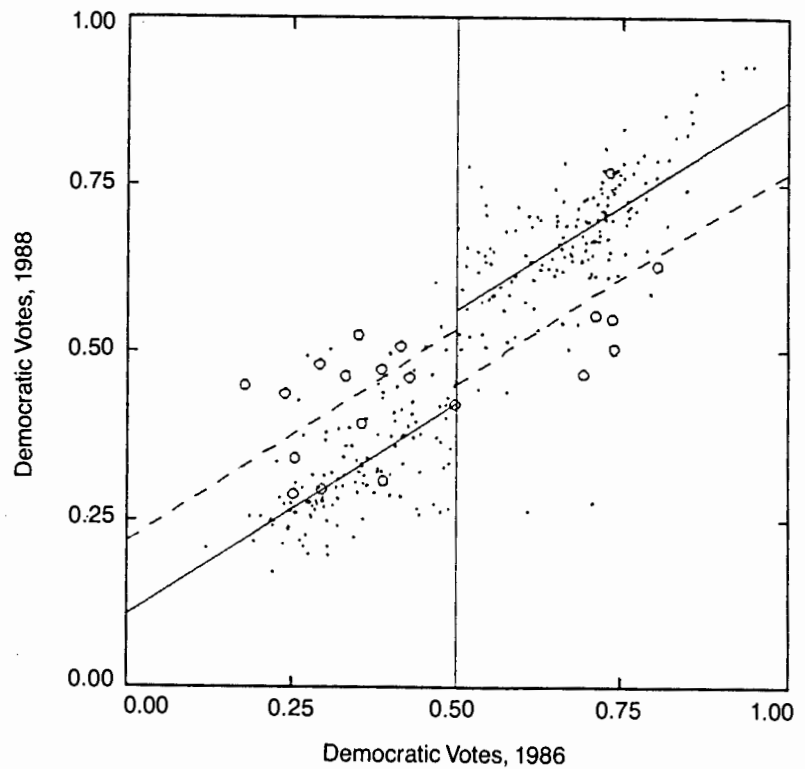


Fig. 6. Electoral swing in contested U.S. House districts. Dots are incumbents and circles are open seats, 1988.

means that of the .043 observed (or total) variation, only a small portion is due to just stochastic variation, leaving the rest as systematic (or true) variation in the unobserved incumbency advantage over the years.

Note that nothing in our methodology constrains the total variance to be more than the estimation variance. However, since true variance can only be zero or positive, it only makes sense for total variance to be at least as large as the estimation variance. If it is not as large, you should question your statistical model, computer program, data, or logic of analysis. The advantage of concentrating on measures of variability like this is that we have a sense of just how much unknown systematic (or true) variation in incumbency advantage exists. Furthermore, it is interesting to note that estimation error in the incumbency advantage (.012) is considerably less than the estimation error in the estimate of the average district-level vote proportions ($\hat{\sigma} = .069$).

We can also go one step further by running a simple time-series regres-

sion of incumbency advantage, calculated for each pair of years, on a linear-trend variable (the election year). The stochastic variability left from this regression is $\hat{\sigma} = .020$, a considerable portion of the total variability of .043.⁹ Thus, aside from the systematic linear increase over time in the incumbency advantage, we still have some remaining stochastic variability. We could add additional variables to this regression to try to explain incumbency advantage over time, but we would not wish for our estimate $\hat{\sigma}$ to drop below the standard error of incumbency advantage of .012, for that would be a clear indication of overfitting. It would make little sense to have a statistical model that purports to predict numbers more precisely than they can be measured. These are also very interesting numbers substantively, of course, since they imply that further research might be conducted into the causes of variation in incumbency advantage over time or across incumbents. Although much research has been conducted about the former, the latter is almost completely unstudied.

All these calculations could be made in terms of R^2 , but it is more difficult and a lot less intuitive. Using R^2 prevents the analyst from speaking directly about politics—in this case, the Democratic proportion of the two-party vote and incumbency advantage—and replaces it with an arbitrary scale from zero to one that has no obvious substantive interpretation. It could be done, and without harm if one is careful, but why do all the work if the end result is a less interpretable product?

6. Concluding Remarks

As Michael Lewis-Beck and Andrew Skalaban imply, close attention to issues of variability can be almost as important as estimates of the effect parameters. We owe them our thanks for opening up this area for discussion. We also owe these issues more serious consideration in our methodological work and empirical research.

REFERENCES

- Edwards, A. W. F. 1972. *Likelihood*. Cambridge: Cambridge University Press.
 Gelman, Andrew, and Gary King. 1990. "Estimating Incumbency Advantage Without Bias." *American Journal of Political Science* 34. Forthcoming.
 King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30:666–87.

9. Measured in variances, the stochastic variance left over from this regression is 22 percent of the observed total variance.

- King, Gary. 1989. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
- King, Gary, and Andrew Gelman. 1991. "Systemic Consequences of Incumbency Advantage in the U.S. House." *American Journal of Political Science*. Forthcoming.
- Lewis-Beck, Michael S. 1980. *Applied Regression: An Introduction*. Beverly Hills: Sage.
- Stigler, Stephen. 1986. *The History of Statistics*. Cambridge, Mass.: Harvard University Press.